Amplify-and-Forward in Wireless Relay Networks

Samar Agnihotri, Sidharth Jaggi, and Minghua Chen
Department of Information Engineering, The Chinese University of Hong Kong
Email: {samar, jaggi, minghua}@ie.cuhk.edu.hk

Abstract

A general class of wireless relay networks with a single source-destination pair is considered. Intermediate nodes in the network employ an amplify-and-forward scheme to relay their input signals. In this case the overall input-output channel from the source via the relays to the destination effectively behaves as an intersymbol interference channel with colored noise. Unlike previous work we formulate the problem of the maximum achievable rate in this setting as an optimization problem with no assumption on the network size, topology, and received signal-to-noise ratio. Previous work considered only scenarios wherein relays use all their power to amplify their received signals. We demonstrate that this may not always maximize the maximal achievable rate in amplify-and-forward relay networks. The proposed formulation allows us to not only recover known results on the performance of the amplify-and-forward schemes for some simple relay networks but also characterize the performance of more complex amplify-and-forward relay networks which cannot be addressed in a straightforward manner using existing approaches.

Using cut-set arguments, we derive simple upper bounds on the capacity of general wireless relay networks. Through various examples, we show that a large class of amplify-and-forward relay networks can achieve rates within a constant factor of these upper bounds asymptotically in network parameters.

I. Introduction

Since their introduction in [1] *Amplify-and-Forward (AF)* relay schemes have been studied in the context of *cooperative communication* [2], [3], estimating the capacity of relay networks [4]–[6], and *analog network coding* [7]–[11]. For cooperative communication, AF schemes provide spatial diversity to fight against fading; for capacity estimation of relay networks, such schemes

provide achievable lower bounds that are known to be optimal in some communication scenarios; and for analog network coding, given the broadcast nature of the wireless medium that allows the *mixing* of the signals in the air, these schemes provide a communication strategy that achieves high throughput with low computational complexity at internal nodes. In this paper, we concern ourselves mostly with the capacity analysis of a general class of Gaussian AF relay networks. Extensions of our method and results to cooperative communication and analog network coding scenarios is part of our future work.

In previous work, while analyzing the performance of AF schemes in relay networks one or more of the following assumptions have been made: networks with a small number of nodes [8], [10]; networks with simple topologies [3]–[5], [8], [10]; or relay operation in the high-SNR regime, [10]. However, for two reasons, we believe that it is important to characterize the performance of the AF schemes without such assumptions. First, we feel that for a scheme such as amplify-and-forward that allows one to exploit the broadcast nature of the wireless medium such assumptions on the size and topology may result in lower achievable performance than otherwise. Second, even in the low-SNR regimes amplify-and-forward can be capacity-achieving relay strategy in some scenarios, [5]. Therefore, a framework to address the performance of AF schemes in general wireless relay networks is desired.

However, one major issue with constructing such a framework is the following. In general wireless relay networks with AF relaying, the resulting input-output channel between the source and the destination is an intersymbol interference (ISI) channel ([4], [10]) with colored noise. This is because both the source signal and the noise introduced at the relay nodes may reach the destination via multiple paths with differing delays. Without the assumptions above, this results in a formidable problem to analyze with the existing methods [4].

Our main contribution is that we provide a framework to compute the maximum achievable rate with AF schemes for a class of general wireless relay networks, namely *Gaussian relay networks*. This framework casts the problem of computing the maximum rate achievable with AF relay networks as an optimization problem. Our work shows that amplifying the received signal to the maximum possible value at intermediate nodes might result in sub-optimal end-to-end throughput. Also, we establish the generality of the proposed formulation by showing that it allows us to derive in a unified and simple manner not only the various existing results on the performance of simple AF relay networks but also new results for more complex networks

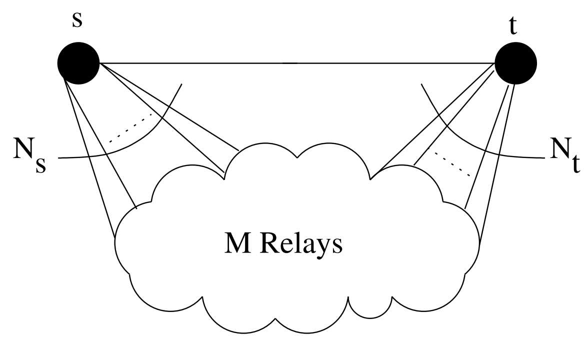

Fig. 1. A single source-single destination communication channel over general Gaussian relay network with M relays.

that cannot be addressed in a straightforward manner with existing methods. We show through various examples that for a large class of relay networks the AF schemes can achieve rates within a constant factor of the cut-set upper-bounds on the capacity of general wireless relay networks.

The paper is organized as follows. In Section II we introduce the general class of Gaussian AF relay networks addressed in this paper. In Section III we formulate the problem of maximum rate achievable via AF schemes in these networks. In Section IV we compute the rates achievable via AF schemes for two instances of such relay networks under various communication scenarios, and then in Section V we discuss the asymptotic behavior of the gap between these rates and the corresponding upper bounds on the capacity of general wireless relay networks computed there. Section VI concludes the paper with a summary.

II. System Model

Let us consider a $(M+2)$-node wireless relay network with source s destination t and M relays as a directed graph $G=(V,E)$ with bidirectional links, as shown in Figure 1. Each node in the network is assumed to have a single antenna. Let us assume that the degree of the source node is N_s+1, with it being connected to the destination node and a subset S_s of the relay nodes, $N_s=|S_s|$. Similarly, let us assume that the degree of the destination node is N_t+1, with it being connected to the source node and a subset S_t of the relay nodes, $N_t=|S_s|$. In general, $S_s \cup S_t \subseteq V \setminus \{s,t\}$.

At instant n, the channel output at node $i, i \in V \setminus \{s\}$, is

$$y_i[n] = \sum_{j \in \mathcal{N}(i)} h_{ji} x_j[n] + z_i[n], \quad -\infty < n < \infty, \tag{1}$$

where $x_j[n]$ is the channel input of the node j in the neighbor set $\mathcal{N}(i)$ of node i. In (1), h_{ji} is a real number representing the channel gain along the link from relay j to relay i. It is assumed to be fixed (for example, as in a single realization of a fading process) and known throughout the network. Further, $\{z_i[n]\}$ is a sequence (in n) of independently and identically distributed (*i.i.d.*) Gaussian random variables with zero mean and variance σ^2, $z_i[n] \sim \mathcal{N}(0, \sigma^2)$. We also assume that z_i are independent of the input signal and of each other. The source symbols $x_s[n], -\infty < n < \infty$, are *i.i.d.* Gaussian random variables with zero mean and variance P_s that satisfies an average source power constraint, $x_s[n] \sim \mathcal{N}(0, P_s)$. We assume that the i^{th} relay node's transmit power is constrained as:

$$E[x_i^2[n]] \leq P_i, \quad -\infty < n < \infty \tag{2}$$

In a general wireless relay network there may exists cycles. If a relay merely amplifies and forwards its received signal in such scenarios then it may be use a significant fraction of its power budget on forwarding the previously forwarded information. Therefore, motivated by some work on analog network coding, such as [12], we propose the following relay operation to allow the relays to expend their transmit power in forwarding only the "new" information.

Relay operation: We assume that each relay node maintains a buffer of signals it forwarded previously. Therefore, each relay node i, after receiving the channel output $y_i[n]$ at time instant n executes the following series of steps:

Step 1: Obtain the *residual signal* $y_i'[n]$ from its input $y_i[n]$ by subtracting the contributions of previously forwarded signals (if any) from $y_i[n]$.

Step 2: Compute the power $P_{R,i}'$ of the *residual signal* $y_i'[n]$.

Step 3: At instant $n+1$ transmit the scaled version of the *residual signal* $y_i'[n]$ of its input at time instant n:

$$x_i[n+1] = \beta_i y_i'[n], \quad 0 \leq \beta_i^2 \leq \beta_{i,max}^2 = P_i / P_{R,i}', \tag{3}$$

where the β_i is the scaling factor[1]. Let the network-wide amplification vector for the M relay

[1]Note that, in general, β_i may depend on i^{th} relay's past observations. $\beta_i[n] = f_{i,n}(Y_i[n-1], \ldots, Y_i[1])$. However, due to practical considerations, such as low-complexity operation, we do not consider such scenarios here.

nodes be denoted as $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$.

Remark 1: One of the major advantages of this relay operation is that by subtracting the previously forwarded signal from its input, a relay node expends its power in forwarding only the "new" information. In a general wireless relay network there may exists cycles. In such scenarios if a relay merely forwards its received signal then it may result in the relay forwarding the scaled version of the linear combination of two or more of its previously forwarded signals. This may provide higher achievable rate as in [13].

Using (1) and (3), the input-output channel between the source and destination can be written as an intersymbol interference (ISI) channel that at instant n is given by

$$\begin{aligned} y_t[n] &= h_{st}x_s[n] + z_t[n] \\ &+ \sum_{d=1}^{D^s}\Bigg[\sum_{(i_1,\ldots,i_d)\in K_d} h_{si_1}\beta_{i_1}h_{i_1i_2}\ldots h_{i_{d-1}i_d}\beta_{i_d}h_{i_dt}\Bigg]x_s[n-d] \\ &+ \sum_{d=1}^{D^1}\Bigg[\sum_{(i_1,\ldots,i_d)\in K_{1,d}} \beta_1 h_{1i_1}\ldots h_{i_{d-1}i_d}\beta_{i_d}h_{i_dt}\Bigg]z_1[n-d] \\ &\;\vdots \\ &+ \sum_{d=1}^{D^M}\Bigg[\sum_{(i_1,\ldots,i_d)\in K_{M,d}} \beta_M h_{Mi_1}\ldots h_{i_{d-1}i_d}\beta_{i_d}h_{i_dt}\Bigg]z_M[n-d], \end{aligned} \tag{4}$$

where $K_d, 1 \leq d \leq D^s$, is the set of d-tuples of node indices corresponding to all paths from the source to the destination with path delay d and D^s is the length of the longest such path. Note that along such paths $D^s \leq M$. Similarly, $K_{m,d}, 1 \leq m \leq M, 1 \leq d \leq D^m$, is the set of d-tuples of node indices corresponding to all paths from the m^{th} relay to the destination with path delay d, D^m is the length of the longest such path from m^{th} relay to the destination. It should be noted that $\max(D^1, \ldots, D^M) = D^s - 1$.

Let us introduce *modified* channel gains as follows. For all the paths between the source s and the destination t:

$$\begin{aligned} &h_0 = h_{st} \\ &h_d = \sum_{(i_1,\ldots,i_d)\in K_d} h_{si_1}\beta_{i_1}h_{i_1i_2}\ldots h_{i_{d-1}i_d}\beta_{i_d}h_{i_dt}, 1 \leq d \leq D^s \end{aligned} \tag{5}$$

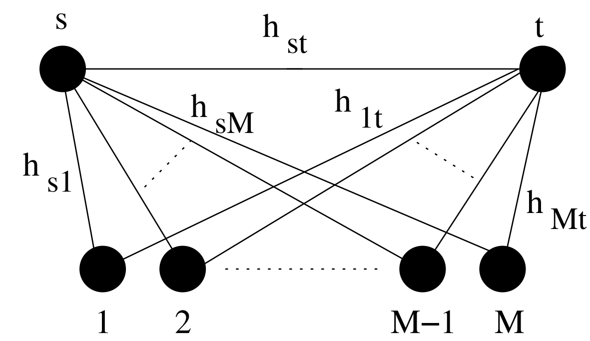

Fig. 2. Type A Relay Network.

For all the paths between the m^{th}-relay, $1 \leq m \leq M$, and the destination t:

$$\begin{aligned} h_{m,0} &= 0, \\ h_{m,d} &= \sum_{(i_1,\ldots,i_d)\in K_{m,d}} \beta_m h_{mi_1} \ldots h_{i_{d-1}i_d} \beta_{i_d} h_{i_d t}, 1 \leq d \leq D^m \end{aligned} \tag{6}$$

Remark 2: Note that though there may be exponentially large number of paths between the source and the destination as well as between a relay and the destination, the modified channel gains for all such paths as in (5) and (6) can be efficiently computed using the line-graphs [14].

In terms of these modified channel parameters, the source-destination ISI channel in (4) can be written as:

$$y_t[n] = \sum_{j=0}^{D^s} h_j x_s[n-j] + \sum_{j=0}^{D^1} h_{1,j} z_1[n-j] + \ldots + \sum_{j=0}^{D^M} h_{M,j} z_M[n-j] + z_t[n] \tag{7}$$

Before proceeding further, let us introduce two special cases of the general class of relay networks introduced in the beginning of this section. In the rest of this paper, we illustrate various concepts and derive some results using these two special networks.

Type A Network: For one source-destination pair and M relays, Type A network is defined as: $G = (V, E)$, where $V = \{s, t, 1, \ldots, M\}$ and $E = \{(s,t), (s,i), (i,t) : i \in \{1, \ldots, M\}\}$. In other words, the source node shares an edge with the destination node and each of M relay nodes. Similarly, the destination node shares an edge with the source node and each of M relay nodes. However, no pair of relay nodes share an edge between themselves. Figure 2 illustrates an instance of Type A networks.

Type B Network: For one source-destination pair and M relays, Type B networks are defined as: $G = (V, E)$, where $V = \{s, t, 1, \ldots, M\}$ and $E = \{(s,t), (s,i), (i,t), (j, j+1) : i \in$

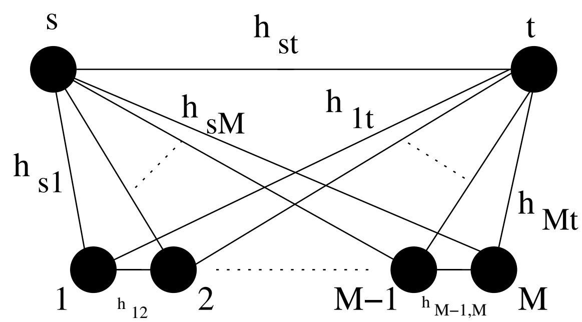

Fig. 3. Type B Relay Network.

$\{1,\ldots,M\}, j \in \{1,\ldots,M-1\}\}$. In other words, as in Type-A networks, the source node shares an edge with the destination node and each of M relay nodes. Similarly, the destination node shares an edge with the source node and each of M relay nodes. However, j^{th} relay shares an edge with $j+1^{\text{st}}$ node, for all $j \in \{1,\ldots,M-1\}$. Figure 3 illustrates an instance of Type B networks.

III. Achievable rates for the source-destination ISI channel in general AF Relay Networks

We first derive the expression for the achievable rate for the source-destination channel in (7) for a given amplification-vector $\boldsymbol{\beta}$ and then formulate the problem of maximizing the achievable rate over the domain of feasible $\boldsymbol{\beta}$.

***Lemma** 1:* For given length-M vector $\boldsymbol{\beta}$, the achievable rate for the channel in (7) with *i.i.d.* Gaussian input is:

$$I(P_s,\boldsymbol{\beta}) = \frac{1}{2\pi}\int_0^{\pi} \log\left[1+\frac{P_s}{\sigma^2}\frac{|H(\lambda)|^2}{1+\sum_{m=1}^{M}|H_m(\lambda)|^2}\right]d\lambda, \tag{8}$$

where

$$H(\lambda) = \sum_{j=0}^{D^s} h_j e^{-ij\lambda},\; H_m(\lambda) = \sum_{j=0}^{D^m} h_{m,j} e^{-ij\lambda},\; i=\sqrt{-1} \tag{9}$$

Proof: In [15] a *Discrete Fourier Transform (DFT)* based formalism is developed to compute the capacity of Gaussian channel with source intersymbol interference (ISI). We compute the maximum achievable rate for the channel in (7) for a given amplification-vector $\boldsymbol{\beta}$ by generalizing this formalism to also include the ISI channel for the Gaussian noise at each relay node resulting

in colored Gaussian noise at the destination[2]. The details of the proof are in the Appendix A. ■

Remark 3: The derivation of an expression for $I(P_s, \boldsymbol{\beta})$ with jointly Gaussian inputs is similar to the proof of Lemma 1. However as such an expression does not aid in the presentation of our ideas we do not discuss it in this paper.

For a given network-wide amplification vector $\boldsymbol{\beta}$, the achievable information rate is given by $I(P_s, \boldsymbol{\beta})$. Therefore the maximum information-rate $I_{AF}(P_s)$ achievable in an AF relay network with *i.i.d.* Gaussian input is defined as the maximum of $I(P_s, \boldsymbol{\beta})$ over all feasible $\boldsymbol{\beta}$, subject to per relay-node amplification constraint (3). In other words:

$$\text{(P1):} \qquad I_{AF}(P_s) \stackrel{def}{=} \max_{\boldsymbol{\beta}: 0 \leq \beta_i^2 \leq \beta_{i,max}^2} I(P_s, \boldsymbol{\beta}) \tag{10}$$

Substituting for $H(\lambda)$ and $H_m(\lambda)$ from (9) in (8), we can rewrite problem (10) equivalently as:

$$\begin{aligned} \text{(P2): } I_{AF}(P_s) &= \max_{\boldsymbol{\beta}: 0 \leq \beta_i^2 \leq \beta_{i,max}^2} I(P_s, \boldsymbol{\beta}) \\ I(P_s, \boldsymbol{\beta}) &= \frac{1}{2\pi} \int_0^{\pi} \log \left[1 + \frac{P_s}{\sigma^2} \frac{\sum_{i=0}^{D^s} A_i \cos(i\lambda)}{\sum_{i=0}^{D^n} B_i \cos(i\lambda)} \right] d\lambda \end{aligned} \tag{11}$$

The coefficients $A_i, 1 \leq i \leq D^s$, and $B_i, 1 \leq i \leq D^n$, are defined in Appendix B.

The formulation of the problem P2 is illustrated in Appendix C for the Type A and Type B relay networks introduced in the previous section.

In general AF relay networks the simultaneous relay transmissions may interfere and if the relays always amplify the received signals to the maximum possible then it may result in sub-optimal end-to-end throughput. Therefore the scaling factor for each relay must be optimally chosen to maximize the achievable rate. This is emphasized by P1 and its significance is illustrated by the following example.

[2]Note that the capacity of discrete-time Gaussian ISI channel was known prior to [15], for example in [25]–[27]. However, the analysis in these papers, based on the asymptotic properties of Toeplitz forms [28], [29], is not easily amenable to derive the capacity results of the ISI channel with colored Gaussian noise. Therefore, like [30], [31], we too use more accessible *DFT* based formalism developed in [15] to compute the maximum achievable rate for the ISI channel with colored Gaussian noise in (7).

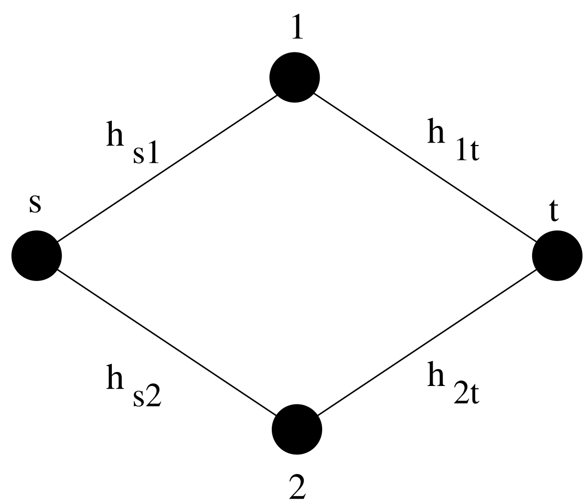

Fig. 4. Diamond network.

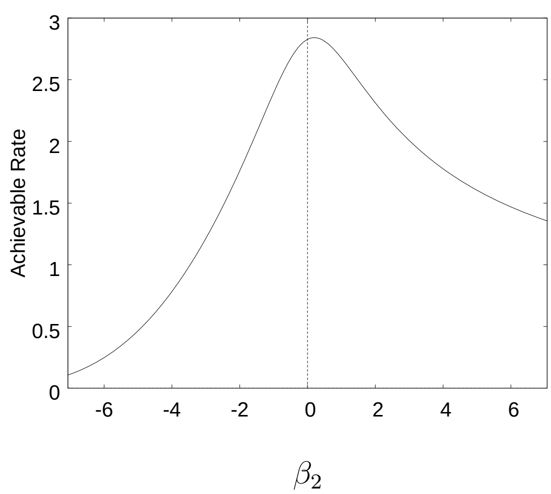

Fig. 5. The achievable rate for the Example 1 when $\beta_1 = \beta_{1,max}$ and β_2 lies in $[-\beta_{2,max}, \beta_{2,max}]$.

Example 1: Let us consider the relay network in Figure 4. Let $h_{s1} = 1, h_{s2} = 0.1, h_{1t} = h_{2t} = 1$. Let $P_s = P_1 = P_2 = 10$ and noise variance $\sigma^2 = 0.1$ at each node. Therefore, we have

$$\beta_{1,max}^2 = \frac{P_1}{h_{s1}^2 P_s + \sigma^2} = 0.99$$
$$\beta_{2,max}^2 = \frac{P_2}{h_{s2}^2 P_s + \sigma^2} = 50.0$$

In this case from (10), we have the following rate maximization problem:

$$I_{AF} = \max_{\beta_1,\beta_2} \frac{1}{2} \log \left[1 + 100 \frac{(\beta_1 + 0.1\beta_2)^2}{1 + \beta_1^2 + \beta_2^2} \right]$$

subject to constraints $0 \leq \beta_1^2 \leq \beta_{1,max}^2$ and $0 \leq \beta_2^2 \leq \beta_{2,max}^2$.

The optimal solution of this problem is $(\beta_1 = 0.995, \beta_2 = 0.225)$. The objective function is plotted in the Figure 5 for $\beta_1 = 0.995$. Therefore, it follows that in this case $\beta_2 = \beta_{2,max}$ is not the optimal amplification factor. ∎

With this observation and the definition of the relay operation given in the previous section, it is appropriate to call the forwarding scheme proposed in this paper as *subtract-scale-and-forward*.

Remark 4 (the computational complexity of the problem P2): Writing the objective function of the problem P2 as sum of ratios, as in (A.14), we can show from [16] that each of these ratios is neither quasiconcave[3] (therefore, more than one local optimum may exist) nor quasiconvex (therefore, the globally optimal solution may not exist at an extreme point or on the boundary of the domain of optimization). The sum of such non-quasiconcave and non-quasiconvex ratios is also non-quasiconcave and non-quasiconvex. Therefore, the problem P2 is a global optimization problem, specifically it belongs to a subset of global optimization problems, called *sum-of-ratios* problems, which are supposed be hard, in general, [17]. Further, in general, $\beta_{i,max}, 1 \leq i \leq M$, depends on $\boldsymbol{\beta}_{-i} \stackrel{def}{=} (\beta_1, \ldots, \beta_{i-1}, \beta_{i+1}, \ldots, \beta_M)$. This dependence of $\beta_{i,max}$ on $\boldsymbol{\beta}_{-i}$ makes the constraint set in (11) non-convex for all but Type A relay networks. Based on these arguments, we conjecture the problem P2 to be computationally hard. However, at present we do not have a formal computational complexity-theoretic proof and constructing such a proof is the part of our future work.

Remark 5 (the approximation schemes to solve the problem P2): The lack of the exact complexity classification of the optimization problem in (11) notwithstanding, we envision solving this problem to be an important step towards characterizing the maximum rates achievable with various other relaying schemes in general wireless relay networks as such schemes also result in problem formulations similar to (11). In [32], we concern ourselves with developing a unified framework to efficiently approximate the optimization problems such as (11).

[3]A function $f : S \to R$ defined on a convex subset S of real vector space is quasiconcave if whenever $x, y \in S$ and $\lambda \in [0, 1]$, then $f(\lambda x + (1 - \lambda)y) \geq \min(f(x), f(y))$, [16].

In the following we show that the problem formulation above not only allows us to reproduce existing results on the achievable rates for some special classes of amplify-and-forward wireless relay networks in a simple manner, but also allows us to compute the achievable rates for much broader class of amplify-and-forward wireless relay networks which could not be hitherto addressed with existing methods.

IV. Approximating the Maximum Achievable Rate $I_{AF}(P_s)$ for Amplify-and-Forward Relay Networks

Let us consider the problem P2 when the relays operate instantaneously as in [8], that is, the relays amplify-and-forward their input signals without delay. Therefore, for relay node $i, 1 \leq i \leq M$, we have $x_i[n] = \beta_i y_i'[n]$. Note the possible system instability resulting from this assumption is avoided by the relay-operation (buffering and subtracting of the previously forwarded signals) as given in the Section II. With this assumption, P2 reduces to (after setting $\lambda = 0$ in (11) and then integrating):

$$\text{(P3): } I_{AF}(P_s) = \max_{\boldsymbol{\beta}:0 \leq \beta_i^2 \leq \beta_{i,max}^2} \frac{1}{2} \log \left[1 + \frac{P_s}{\sigma^2} \frac{\sum_{i=0}^{D^s} A_i}{\sum_{i=0}^{D^n} B_i} \right] \tag{12}$$

The problem P3 can also be derived directly from the channel model in (7), however we do not discuss that derivation here for the sake of brevity.

Given the particular form of functional dependence of A_is and B_is on the components of the amplification vector $\boldsymbol{\beta}$ (as follows from (4), (5), (B.2), and (B.9)) and dependence of β_is on each other, the problem P3 is a *Geometric Program (GP)* [18], where both, the objective function as well as the constraints are expressed in terms of *posynomials*. However, minimizing or upper bounding (equivalently, maximizing or lower bounding) a ratio of two posynomials is a non-convex problem that is intrinsically intractable [19], [20][4]. Therefore, for a general relay network, it is not possible to exactly solve or lower bound the problem P3 in a computationally efficient manner. However, unlike the problem P2 in (11), for the problem P3, efficient approximation schemes exist that solve such problems iteratively by solving a series of GPs, as discussed in [19,

[4]Note that in [19], [20] the objective function can be written as an inverted posynomial in high-SINR regime. Therefore the problem is efficiently solvable using GP methods. It is only in low to moderate SINR regimes where the objective function cannot be so written, the problem is NP-hard. However, the objective function in (12) involves the ratio of posynomials for all SNR values. Therefore, in general, the problem is hard irrespective of SNR value.

Section 3.3]. Also, for some specific relay networks, under some assumptions, we can efficiently compute the lower bounds, as we show next.

Let us consider the problem P3 when the relay nodes are constrained to use the same amplification-factor, that is, $\beta_i = \beta$, for all $1 \leq i \leq M$. In the practical setting, this assumption considerably simplifies the system-design with β set to one particular value for all relay nodes. Then, (P3) reduces to

$$\text{(P4):} \qquad I_{AF}(P_s) = \max_{0 \leq \beta^2 \leq \beta^2_{max}} I_{AF}(P_s, \beta) \tag{13}$$

$$I_{AF}(P_s, \beta) = \frac{1}{2} \log \left[1 + \frac{P_s}{\sigma^2} \frac{\sum_{i=0}^{D^s} A_i}{\sum_{i=0}^{D^n} B_i} \right] \tag{14}$$

Note that the solution of (P3) cannot be smaller than the solution of (P4) because the set of feasible $\boldsymbol{\beta} : \beta_i = \beta$, for (P4) is a subset of the set of feasible $\boldsymbol{\beta}$ for (P3).

Note 1: In general, $\beta^2_{i,max} \sim P_i$. Therefore $\sum \beta^2_{i,max} \sim \sum P_i$. However, as we are considering the case of equal β, $\beta_{i,max} = \beta_{max}$, so we have $M\beta^2_{max} \sim \sum P_i$ or $\beta^2_{max} \sim M^{-1} \sum P_i$.

Now we discuss solving the problem P4 for Type-A and Type-B relay networks introduced in Section II, in different communication scenarios. The proofs of various propositions and lemmas are provided in the Appendix D.

A. *Type-A Relay Network*

Let us first consider Type-A relay network as in Figure 2. For such networks, we solve the problem P4 in the following two scenarios.

Scenario 1 (No attenuation network): Let us assume that there is no attenuation along any link in the network, that is, $h_{st} = h_{si} = h_{it} = 1$ for all $1 \leq i \leq M$. The problem P4 in this case is:

Proposition *1:*

$$I_{AF}(P_s) = \max_{0 \leq \beta^2 \leq \beta^2_{max}} \frac{1}{2} \log \left[1 + \frac{P_s}{\sigma^2} \frac{(1 + M\beta)^2}{1 + M\beta^2} \right]$$

Lemma *2:* $I_{AF}(P_s)$ attains its global maximum at $\beta_{opt} = 1$.

Now let us consider two particular ways in which β_{max} varies with network size.

Scenario 1, Case A (Increasing relay power): For the given network with M relay nodes, let us assume that the sum power of the relay nodes is constrained as follows:

$$\sum_{m=1}^{M} E[X_m^2] \leq \sum_{m=1}^{M} P_i \leq M^{u+1} Q, u > 0, Q = constant$$

So $\beta^2_{max} = M^u Q$. From Lemma 2, we have for $M \to \infty$.

$$I_{AF}(P_s) = \begin{cases} \frac{1}{2}\log[1+\frac{P_s}{\sigma^2}(1+M)], \text{ if } \beta_{max} \geq 1, \\ \\ \frac{1}{2}\log[1+\frac{P_s}{\sigma^2}(M+\frac{2}{\beta_{max}})], \text{ otherwise} \end{cases} \tag{15}$$

Scenario 1, Case B (Constant total relay power): Let us consider the case where the sum power of relay nodes is fixed irrespective of the number of relay nodes, that is $\sum_{m=1}^{M} P_i \leq Q, Q = constant$. Therefore, we set $\beta^2_{max} = \frac{Q}{M}$. As $M \to \infty$, for sufficiently large M, $\beta_{max} < \beta_{opt} = 1$. Therefore, from Lemma 2, $\beta = \beta_{max}$ maximizes the achievable rate and we have for $M \to \infty$

$$I_{AF}(P_s) = \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{Q}{1+Q}M\right] \tag{16}$$

Scenario 2 (Bounded channel gains): In Scenario 1, we considered no attenuation relay networks. Now, let us consider the scenario where the channel gains are arbitrary, but strictly bounded, $0 < h_{st}, h_{si}, h_{it} < \infty$, $1 \leq i \leq M$. The problem P4 in this case is:

***Proposition** 2:*

$$I_{AF}(P_s) = \max_{0 \leq \beta^2 \leq \beta^2_{max}} \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{(h_{st}+\beta\sum_{i=1}^{M}h_{si}h_{it})^2}{1+\beta^2\sum_{i=1}^{M}h_{it}^2}\right]$$

***Lemma** 3:* $I_{AF}(P_s)$ attains its global maximum at $\beta_{opt} = \frac{\sum_{i=1}^{M}h_{si}h_{it}}{h_{st}\sum_{i=1}^{M}h_{it}^2}$.

Increasing relay power: Let us consider the increasing total relay power scenario as in Scenario 1, Case A. Let $\beta^2_{max} = M^u Q$. In this case, following Lemma 3, we obtain the following lower bound on the achievable rate as $M \to \infty$:

$$\begin{aligned} I_{AF}(P_s) &> \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}(Mh_{min}+1)\frac{h^2_{s,max}h_{min}}{h_{max}}\right], \\ &\quad \text{if } \beta_{max} \geq \frac{\sum_{i=1}^{M}h_{si}h_{it}}{h_{st}\sum_{i=1}^{M}h_{it}^2}, \\ I_{AF}(P_s) &> \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}(Mh_{min}+\frac{1}{\beta_{max}})\frac{h^2_{s,max}h_{min}}{h_{max}}\right], \\ &\quad \text{otherwise,} \end{aligned} \tag{17}$$

where $h_{s,max} = \max\{h_{st}, h_{s1}, \ldots, h_{sM}\}$, $h_{min} = \min\{h_{1t}, \ldots, h_{Mt}\}$, and $h_{max} = \max\{h_{1t}, \ldots, h_{Mt}\}$.

B. Type-B Relay Network

Let us consider the Type-B relay network as in Figure 3. For such networks, we consider the no-attenuation scenario where all channel gains are set to unity, that is, $h_{st} = h_{si} = h_{it} = 1$, $1 \leq i \leq M$ as well as $h_{i,i+1} = 1$, $1 \leq i \leq M-1$. The problem P4 in this case is:

***Proposition** 3:*

$$
\begin{aligned}
I_{AF}(P_s,\beta) &= \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{\left(1+\beta M\frac{\beta^M-1}{\beta-1}-\frac{M\beta^{M+1}}{\beta-1}+\beta^2\frac{\beta^M-1}{(\beta-1)^2}\right)^2}{1+\sum_{i=1}^{M}\beta^2\left(\frac{\beta^{M-(i-1)}-1}{\beta-1}\right)^2}\right] \\
&= \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{\left(1+\beta M\frac{\beta^M-1}{\beta-1}-\frac{M\beta^{M+1}}{\beta-1}+\beta^2\frac{\beta^M-1}{(\beta-1)^2}\right)^2}{1+\frac{\beta^2}{(\beta-1)^2}\left[\beta^2\frac{\beta^{2M}-1}{\beta^2-1}-2\beta\frac{\beta^M-1}{\beta-1}+M\right]}\right] \\
I_{AF}(P_s) &= \max_{0\leq\beta\leq\beta_{max}} I_{AF}(P_s,\beta)
\end{aligned}
\tag{18}
$$

It can be proved that the objective function is quasiconcave, therefore a unique global maximum exists. Let β that solves (18) be denoted as $\beta = \beta_{opt}$. However, obtaining a closed-form expression for β_{opt} does not appear straightforward, though it can be numerically computed for any M.

Constant total relay power: Let us consider the case where the sum power of relay nodes is fixed irrespective of the number of relay nodes as in Scenario 1, Case B. Let $\beta_{max}^2 = \frac{Q}{M}$. As $M \to \infty$, for sufficiently large M, $\beta_{max} < \beta_{opt}$. Therefore, $\beta = \beta_{max}$ maximizes the achievable rate and we have the following rate achievable asymptotically as $M \to \infty$

$$
I_{AF}(P_s) = \frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{QM}{1+Q}\frac{1+\sqrt{Q/M}}{1-\sqrt{Q/M}}\right] \tag{19}
$$

V. Asymptotic Capacity

In the previous two sections, we formulated the problem of maximum achievable rate for AF relay networks and then we computed explicit lower bounds to the capacity of two specific AF relay networks in various communication scenarios. In this section, we first derive an upper bound to the capacity of the general relay networks we address in this paper as introduced in the Section II. We then discuss the asymptotic behavior of the gap between this upper bound and the lower bounds computed for two specific relay networks in the previous section.

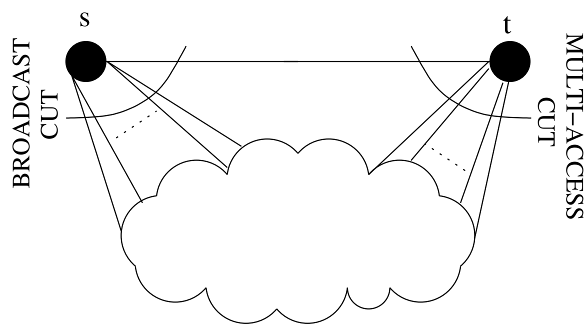

Fig. 6. Illustration of the particular broadcast and multiple-access cuts of the Gaussian relay network in Figure 1.

A. Upper bounds to Capacity of Relay Networks

In [4], an upper bound to the capacity of the relay network of Type A is computed using a weaker corollary of the cut-set theorem [22, Theorem 15.10.1] and the capacity formula for Gaussian vector channels with fixed transfer function [23]. Using this corollary and the capacity formula, we can also compute the upper bound to the capacity of the general relay network introduced in the Section II, as stated in the following proposition.

Proposition *4:* The capacity C of a general wireless relay network is upper-bounded as $C \leq \min\{C_{BC}, C_{MAC}\}$, where C_{BC} and C_{MAC} are the upper bounds on the capacity of the broadcast cut and multiple-access cut respectively, as in the Figure 6, and are given as follows

$$\begin{aligned} C_{BC} &= \log\left[1 + \frac{P_s}{\sigma^2}(h_{st}^2 + \sum_{i \in S_s} h_{si}^2)\right] \\ C_{MAC} &= \log\left[1 + \frac{P_s + \sum_{i \in S_t} P_i}{\sigma^2}(h_{st}^2 + \sum_{i \in S_t} h_{it}^2)\right] \end{aligned}$$

Proof: The details of the proof are provided in the Appendix E. ■

Remark 6: Proposition 3 in [4] can be obtained as a special case of the above proposition by setting $N_s = N_t = M$.

B. Type A Relay Networks

No attenuation, increasing relay power: In this case, the broadcast bound C_{BC} is always asymptotically smaller than the multiple-access bound C_{MAC}, as follows from

$$2^{C_{MAC} - C_{BC}} \approx 1 + \frac{M^{u+1}Q}{P}, \text{ for large } M.$$

Therefore it suffices to compute the asymptotic gap between C_{BC} and the lower bound in (15). In fact, in this case we have

$$C_{BC} - 2I_{AF}(P_s) = 0, \text{ for all } M \geq 1$$

The actual capacity C of the relay network in this case is bounded by $C_{BC}/2 \leq C \leq C_{BC}$.

No attenuation, constant total relay power: In this case also the broadcast bound C_{BC} is asymptotically smaller than the multiple access-bound C_{MAC}, as shown below

$$\lim_{M\to\infty} 2^{C_{MAC}-C_{BC}} = 1 + \frac{Q}{P}$$

Therefore we only address the asymptotic gap between C_{BC} and the lower bound in (16). We have

$$\lim_{M\to\infty} C_{BC} - 2I_{AF}(P_s) = \frac{1}{2}\log(1 + 1/Q)$$

The actual capacity C of the relay network in this case is bounded by $\frac{1}{2}(C_{BC} - \frac{1}{2}\log(1+1/Q)) \leq C \leq C_{BC}$ and the bound gets tighter with increasing Q.

Bounded channel gains: The gap between C_{BC} and lower bound of achievable rate in (17) is bounded asymptotically as:

$$\lim_{M\to\infty} C_{BC} - 2I_{AF}(P_s) \leq \frac{1}{2}\log\left[\frac{h_{max}}{h_{min}^2}\right]$$

The apparent looseness of the gaps computed above compared to the corresponding gaps in [4] arises from the series of simplifications made to reduce the problem (P2) to the problem (P4) and the particular definition of lower bound used in [4].

C. Type B Relay Networks

No attenuation, constant total relay power: As claimed above, the upper-bound in Proposition 4 holds for Type B networks too. Therefore we only address the asymptotic gap between C_{BC} and the lower bound in (19). We have

$$\lim_{M\to\infty} C_{BC} - 2I_{AF}(P_s) = \frac{1}{2}\log(1 + 1/Q)$$

The actual capacity C of the relay network in this case too is bounded by $\frac{1}{2}(C_{BC} - \frac{1}{2}\log(1 + 1/Q)) \leq C \leq C_{BC}$ and bound gets tighter with increasing Q.

VI. Conclusion and Future Work

We provide a framework to analyze the performance of the AF relay schemes in a general class of wireless relay networks. We demonstrate the effectiveness of the proposed framework in two ways. We first show that some well-established results for Gaussian relay networks (for example, those in [4]) can be derived in a much simpler manner. Then, we show that the achievable rates for more complex Gaussian relay networks, which cannot be addressed with existing methods, can be derived straightforwardly. The proposed framework allows for the computation of tighter amplify-and-forward lower bounds to the capacity of Gaussian relay networks.

Future Work: We have provided the analysis of AF scheme for different scenarios with the assumption of instantaneous relay operation. However, the resulting performance is not optimal when delay in relay operation is included, as the constant power allocation is not optimal for the original problem (11). We also want to formally prove the hardness of optimization problem (11) and develop efficient approximation schemes for it. An extension of our work also facilitates the computation of achievable rates for analog network coding scenarios for non-layered networks and in low to moderate SNR regimes that cannot be addressed with existing approaches. We plan to address it in detail in our future work.

Acknowldegment

This work has been supported by the CERG grant 412207.

References

[1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Trans. Inform. Theory,* vol. IT-50, December 2004.

[2] Y. Zhao, R. Adve, and T. J. Lim, "Improving amplify-and-forward relay networks: optimal power allocation versus selection," *IEEE Trans. Wireless. Comm.,* vol. TWC-6, August 2007.

[3] S. Borade, L. Zheng, and R. Gallager, "Amplify-and-forward in wireless relay networks: Rate, diversity, and network size," *IEEE Trans. Inform. Theory,* vol. IT-53, October 2007.

[4] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inform. Theory,* vol. IT-51, March 2005.

[5] K. S. Gomadam and S. A. Jafar, "Optimal relay functionality for SNR maximization in memoryless relay networks," *IEEE JSAC,* vol. 25, February 2007.

[6] T. Cui, T. Ho, and J. Kliewer, "Memoryless relay strategies for two-way relay channels," *IEEE Trans. Comm.,* vol. 57, October 2009.

[7] S. Zhang, S. C. Liew, and P. P. Lam, “Physical-Layer Network Coding,” *Proc. ACM MobiCom,* Los Angeles, CA, September 2006.

[8] S. Katti, I. Marić, A. Goldsmith, D. Katabi, and M. Médard, “Joint relaying and network coding in wireless networks,” *Proc. IEEE ISIT 2007,* Nice, France, June 2007.

[9] S. Katti, S. Gollakotta, and D. Katabi, “Embracing wireless interference: analog network coding,” *Proc. ACM SIGCOMM,* Kyoto, Japan, August, 2007.

[10] I. Marić, A. Goldsmith, and M. Médard, “Analog network coding in the high-SNR regime,” *Proc. IEEE WiNC 2010,* Boston, MA, June 2010.

[11] A. Argyriou and A. Pandharipande, “Cooperative protocol for analog network coding in distributed wireless networks,” *IEEE Trans. Wireless. Comm.,* vol. TWC-9, October 2010.

[12] Q. You, Z. Chen, Y. Li, and B. Vucetic, “Multi-hop bi-directional relay transmission schemes using amplify-and-forward and analog network coding,” *To appear in Proc. IEEE ICC 2011,* Kyoto, Japan, June 2011.

[13] S. Zahedi, M. Mohseni, and A. El Gamal, “On the capacity of AWGN relay channels with linear relaying functions,” ,” *Proc. IEEE ISIT 2004,* Chicago, IL, June 2004.

[14] R. Koetter and M. Médard, “An algebraic approach to network coding,” *IEEE/ACM Trans. Netw.,* vol. 11, October 2003.

[15] W. Hirt and J. L. Massey, “Capacity of the discrete-time Gaussian channel with intersymbol interference,” *IEEE Trans. Inform. Theory,* vol. IT-34, May 1988.

[16] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang, *Generalized Concavity*. Plenum Press, 1988.

[17] S. Schaible and J. Shi, “Fractional Programming: the sum-of-ratios case,” *Optimization Methods and Software,* vol. 18, April 2003.

[18] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, “A tutorial on Geometric Programming,” *Optimization and Engineering,* vol. 8, March 2007.

[19] M. Chiang, “Geometric Programming for communication systems,” Foundations and Trends in Communications and Information Theory, vol. 2, August 2005.

[20] M. Chiang, C. W. Tan, D. P. Palomar, D. O’Neill, and D. Julian, “Power control by Geometric Programming,” *IEEE Trans. Wireless. Comm.,* vol. TWC-6, July 2007.

[21] M. J. Osborne, http://www.economics.utoronto.ca/osborne/MathTutorial/QCC.HTM

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2/e, John Wiley & Sons, 2006.

[23] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Europ. Trans. Telecommun.,* vol. 10, Nov./Dec. 1999.

[24] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[25] B. S. Tsybakov, “Capacity of a discrete-time Gaussian channel with a filter,” *Probl. Peredach. Inform.,* vol. 6, pp. 78-82, 1970.

[26] W. Toms and T. Berger, “Capacity and error exponents of a channel modeled as a linear dynamic system,” *IEEE Trans. Inform. Theory,* vol. IT-19, January 1973.

[27] L. H. Brandenburg and A. D. Wyner, “Capacity of the Guassian channel with memory: The multivariate case,” *Bell Syst. Tech. J.,* vol. 53, May-June 1974.

[28] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. 2/e, Chelsea Publishing Co., New York, 1984.

[29] R. M. Gray, “On the asymptotic eigenvalue distribution of Toeplitz matrices,” *IEEE Trans. Inform. Theory,* vol. IT-18, November 1972.

[30] A. J. Goldsmith and M. Effros, "The capacity region of broadcast channels with intersymbol interference and colored Gaussian noise," *IEEE Trans. Inform. Theory,* vol. IT-47, January 2001.

[31] C. Choudhuri and U. Mitra, "Capacity of relay Channels with ISI and colored Gaussian noise," *Proc. IEEE ISIT 2009,* Seoul, Korea, June 2009.

[32] S. Agnihotri, M. Chen, and S. Jaggi, "Approximation schemes for a particular class of global optimization problems encountered in the capacity analysis of the ISI channels with colored noise," *Under preparation,* 2011.

APPENDIX A
PROOF OF LEMMA 1

Following Gallager [24], we define the information rate for the channel in (7) as

$$I(P_s, \boldsymbol{\beta}) = \lim_{N\to\infty} I_N(P_s, \boldsymbol{\beta}), \tag{A.1}$$

where

$$I_N(P_s, \boldsymbol{\beta}) = N^{-1} I(x_s[0, N-1]; y_t[0, N-1]), \tag{A.2}$$

with the assumption

$$x[-D^s, -1] = (0, \dots, 0)$$

It should be noted that as D^s is finite, this assumption does not affect $I(P_s, \boldsymbol{\beta})$, but it simplifies the analysis as we show later.

Following [15], we introduce a new channel model whose maximum achievable rate can be easily computed. The capacity of this channel can be, subsequently, used to compute $I(P_s)$. The output of the new channel at time instant n is given as

$$\begin{aligned}\tilde{y}_t[n] \; = \; & \sum_{j=0}^{N-1} \tilde{h}_j x_s[(n-j) \bmod N] \\ & + \sum_{j=0}^{N-1} \tilde{h}_{1,j} z_1[(n-j) \bmod N] \\ & + \ldots + \sum_{j=0}^{N-1} \tilde{h}_{M,j} z_M[(n-j) \bmod N] \\ & + z_t[n], \quad 0 \le n < N,\end{aligned} \tag{A.3}$$

where $N > D^s$. Extending the unit-sample response of the channel for signal $h[0, D^S]$ with $N - D^s - 1$ zeros, we define $\tilde{h}[0, N-1] = (h_0, h_1, \dots, h_{D^s}, 0, 0, \dots, 0)$. We can similarly extend the corresponding unit-sample responses of the channel for noise samples $z_1, z_2, \dots, z_M$ and define

$\tilde{h}_1[0,N-1] = (h_{1,0}, h_{1,1}, \ldots, h_{1,D^1}, 0, 0, \ldots, 0)$, $\tilde{h}_2[0,N-1] = (h_{2,0}, h_{2,1}, \ldots, h_{2,D^2}, 0, 0, \ldots, 0)$, $\ldots$, $\tilde{h}_M[0,N-1] = (h_{M,0}, h_{M,1}, \ldots, h_{M,D^M}, 0, 0, \ldots, 0)$, respectively. With the definitions of these extended sample responses, we can rewrite (A.3) symbolically as

$$\begin{aligned} \tilde{y}_t[0,N-1] &= x_s[0,N-1] \circledast \tilde{h}[0,N-1] \\ &\quad + z_1[0,N-1] \circledast \tilde{h}_1[0,N-1] \\ &\quad + \ldots + z_M[0,N-1] \circledast \tilde{h}_M[0,N-1] \\ &\quad + z_t[0,N-1], \end{aligned} \tag{A.4}$$

where $\circledast$ denotes the circular convolution operator. The transmit power constraint (2) leads to the power constraint

$$E[x_i^2[n]] \leq P_i, \quad 0 \leq n < N, \tag{A.5}$$

for the new channel model.

The information rate for the channel defined by (A.3) and (A.5) is defined as

$$\tilde{I}_N(P_s, \boldsymbol{\beta}) = N^{-1} I(x_s[0,N-1]; \tilde{y}_t[0,N-1]) \tag{A.6}$$

Taking the *Discrete Fourier Transform (DFT)* on both sides of (A.4) we have

$$\tilde{Y}_i = \tilde{H}_i X_i + \tilde{H}_{1,i} Z_{1,i} + \ldots + \tilde{H}_{M,i} Z_{M,i} + Z_i, \quad 0 \leq i < N, \tag{A.7}$$

where $\tilde{Y}_i$, $\tilde{H}_i$, X_i, $\tilde{H}_{1,i}$, $Z_{1,i}$, $\ldots$, $\tilde{H}_{M,i}$, $Z_{M,i}$, and Z_i are the components of $DFT\{\tilde{y}_t[0,N-1]\}$, $DFT\{\tilde{h}[0,N-1]\}$, $DFT\{x_s[0,N-1]\}$, $DFT\{\tilde{h}_1[0,N-1]\}$, $DFT\{z_1[0,N-1]\}$, $\ldots$, $DFT\{\tilde{h}_M[0,N-1]\}$, and $DFT\{z_M[0,N-1]\}$, respectively.

Dividing both sides of this by $\tilde{H}_i$ and transforming the resulting first $\lfloor N/2 \rfloor + 1$ components with the transformation [15, eq. (24)], we get the following equivalent form of (A.3) in the transform domain:

$$Y'_i = X'_i + Z'_{1,i} + \ldots + Z'_{M,i} + Z'_i, \quad 0 \leq i < N \tag{A.8}$$

where X'_i, $Z'_{1,i}$, $\ldots$, $Z'_{M,i}$, Z'_i are obtained from transform [15, eq. (24)] with $B_i \equiv X_i$, $B_i \equiv \tilde{H}_{1,i} Z_{1,i} / \tilde{H}_i$, $\ldots$, $B_i \equiv \tilde{H}_{M,i} Z_{M,i} / \tilde{H}_i$, and $B_i \equiv Z_i / \tilde{H}_i$, respectively.

Note that

$$\begin{aligned}
\tilde{H}_{1,i}/\tilde{H}_i &= \tilde{H}^*_{1,N-i}/\tilde{H}^*_{N-i}, \quad 0 \le i < N \\
&\vdots \\
\tilde{H}_{M,i}/\tilde{H}_i &= \tilde{H}^*_{M,N-i}/\tilde{H}^*_{N-i}, \quad 0 \le i < N \\
1/\tilde{H}_i &= 1/\tilde{H}^*_{N-i}, \quad 0 \le i < N
\end{aligned}$$

where $*$ denotes the complex conjugate. Therefore, it follows from [15, Lemma 1] that $Z'_{m,i}, 1 \le m \le M$, in (A.8) are statistically independent zero mean Gaussian random variables with variance

$$\sigma^2_{m,i} = N\sigma^2 \frac{|\tilde{H}_{m,i}|^2}{|\tilde{H}_i|^2}, \quad 0 \le i < N. \tag{A.9}$$

Similarly, Z'_i are statistically independent zero mean Gaussian random variables with variance

$$\sigma^2_i = N\sigma^2 \frac{1}{|\tilde{H}_i|^2}, \quad 0 \le i < N \tag{A.10}$$

Therefore, the equivalent transform domain channel model (A.8) is a set of N parallel discrete memoryless additive Gaussian noise channels with total noise variance

$$\sigma^2_i + \sum_{m=1}^{M} \sigma^2_{m,i} = N\sigma^2 \frac{1 + \sum_{m=1}^{M} |\tilde{H}_{m,i}|^2}{|\tilde{H}_i|^2}, \quad 0 \le i < N \tag{A.11}$$

Further, using [15, Lemma 2], it follows that the transformed inputs $X'_i, 0 \le i, < N$, are i.i.d. zero mean Gaussian random variables with variance NP_s. Therefore, the equivalent transform-domain channel model in (A.8) is a set of N parallel, independent, discrete, memoryless additive Gaussian noise channels with zero mean Gaussian inputs of variance NP_s and total noise variance given in (A.11). This implies that the average mutual information of i^{th} component channel is then the capacity of point-to-point AWGN channel, given by

$$I(X'_i, Y'_i) = \frac{1}{2} \log \left[1 + \frac{P_s}{\sigma^2} \frac{|\tilde{H}_i|^2}{1 + \sum_{m=1}^{M} |\tilde{H}_{m,i}|^2} \right] \tag{A.12}$$

As the N component channels are mutually independent, therefore $\tilde{I}_N(P_s, \boldsymbol{\beta})$ in (A.6) is given by

$$\tilde{I}_N(P_s, \boldsymbol{\beta}) = \frac{1}{2N} \sum_{i=0}^{N-1} \log \left[1 + \frac{P_s}{\sigma^2} \frac{|\tilde{H}_i|^2}{1 + \sum_{m=1}^{M} |\tilde{H}_{m,i}|^2} \right] \tag{A.13}$$

Further, [15, Theorem 4] states that

$$\begin{aligned}(1-D^s/N)I_{N-D^s}(P_s,\boldsymbol{\beta}) &\leq \tilde{I}_N(P_s,\boldsymbol{\beta}) \\ &\leq (1+D^s/N)I_{N+D^s}(P_s,\boldsymbol{\beta})\end{aligned}$$

Therefore, we finally have $I(P_s,\boldsymbol{\beta})$ in (A.1) as

$$\begin{aligned}I(P_s,\boldsymbol{\beta}) &= \lim_{N\to\infty}\tilde{I}_N(P_s,\boldsymbol{\beta}) \qquad \text{(A.14)} \\ &= \lim_{N\to\infty}\frac{1}{2N}\sum_{i=0}^{N-1}\log\left[1+\frac{P_s}{\sigma^2}\frac{|\tilde{H}_i|^2}{1+\sum_{m=1}^{M}|\tilde{H}_{m,i}|^2}\right] \\ &\overset{(a)}{=} \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log\left[1+\frac{P_s}{\sigma^2}\frac{|H(\lambda)|^2}{1+\sum_{m=1}^{M}|H_m(\lambda)|^2}\right]d\lambda \\ &\overset{(b)}{=} \frac{1}{2\pi}\int_{0}^{\pi}\log\left[1+\frac{P_s}{\sigma^2}\frac{|H(\lambda)|^2}{1+\sum_{m=1}^{M}|H_m(\lambda)|^2}\right]d\lambda\end{aligned}$$

where (a) follows from the property of Riemann integrals as stated in [15, Lemma 5] and (b) follows from the fact that $|H(\lambda)| = |H(-\lambda)|, |\lambda| \leq \pi$. It should be noted that $H(\lambda)$ is the transfer function of the filter with unit-sample response $(h_0, h_1, \ldots, h_{D^s})$, given as follows

$$H(\lambda) = \sum_{j=0}^{D^s} h_j e^{-ij\lambda}, \quad i=\sqrt{-1} \qquad \text{(A.15)}$$

Similarly, $H_m(\lambda)$ is the transfer function of the filter with unit-sample response $(h_{m,0}, h_{m,1}, \ldots, h_{m,D^m})$, $1 \leq m \leq M$, given by

$$H_m(\lambda) = \sum_{j=0}^{D^m} h_{m,j} e^{-ij\lambda}, \quad i=\sqrt{-1} \qquad \text{(A.16)}$$

APPENDIX B
EXPRESSIONS FOR $|H(\lambda)|^2$ AND $|H_m(\lambda)|^2$

Following (9), we have

$$\begin{aligned}|H(\lambda)|^2 &= H(\lambda)H^*(\lambda) \\ &= \sum_{j=0}^{D^s} h_j e^{-ij\lambda}\sum_{j=0}^{D^s} h_j e^{ij\lambda} \\ &= \sum_{j=0}^{D^s} h_j^2 + 2(\sum_{j=0}^{D^s-1} h_j h_{j+1})\cos\lambda + 2(\sum_{j=0}^{D^s-2} h_j h_{j+2})\cos(2\lambda) \qquad \text{(B.1)} \\ &\quad + \ldots + 2h_0 h_{D^s}\cos(D^s\lambda) \\ &= A_0 + A_1\cos\lambda + \ldots + A_{D^s}\cos(D^s\lambda) \qquad \text{(B.2)}\end{aligned}$$

Comparing the coefficients of $\cos(j\lambda), 1 \leq j \leq D^s$, in (B.1) and (B.2), defines A_j.

Note that we have the following relation between A_is and h_is:

$$|H(0)|^2 = \sum_{i=0}^{D^s} A_i = (h_0 + \ldots + h_{D^s})^2 \tag{B.3}$$

Noting that $\cos(n\lambda)$ can be expressed in terms of $T_n(x)$, the Chebyshev polynomials of the first kind, as $\cos(n\lambda) = T_n(\cos\lambda)$, we can rewrite $|H(\lambda)|^2$ as

$$\begin{aligned}
|H(\lambda)|^2 &= \sum_{j=0}^{\lfloor D^s/2 \rfloor} (-1)^j A_{2j} + (\sum_{j=0}^{\lfloor D^s/2 \rfloor} (-1)^j (2j+1) A_{2j+1}) \cos\lambda + \ldots + 2^{D^s-1} A_{D^s} \cos^{D^s}\lambda & \text{(B.4)}\\
&= A'_0 + A'_1 \cos\lambda + \ldots + A'_{D^s} \cos^{D^s}\lambda & \text{(B.5)}
\end{aligned}$$

Comparing the coefficients of $\cos^j\lambda, 1 \leq j \leq D^s$, in (B.4) and (B.5), defines A'_j.

Similarly, following (9), the coefficients $A'_{m,j}$ and $A_{m,j}$, $1 \leq j \leq D^m$, are obtained as follows:

$$\begin{aligned}
|H_m(\lambda)|^2 &= H_m(\lambda) H_m^*(\lambda) \\
&= \sum_{j=0}^{D^m} h_{m,j} e^{-ij\lambda} \sum_{j=0}^{D^m} h_{m,j} e^{ij\lambda} \\
&= \sum_{j=0}^{D^m} h_{m,j}^2 + 2(\sum_{j=0}^{D^m-1} h_{m,j} h_{m,j+1}) \cos\lambda + 2(\sum_{j=0}^{D^m-2} h_{m,j} h_{m,j+2}) \cos(2\lambda) \\
&\quad + \ldots + 2 h_{m,0} h_{m,D^m} \cos(D^m\lambda) \\
&\overset{(a)}{=} A_{m,0} + A_{m,1} \cos\lambda + \ldots + A_{m,D^m} \cos(D^m\lambda) & \text{(B.6)}\\
&= \sum_{j=0}^{\lfloor D^m/2 \rfloor} (-1)^j A_{2j} + (\sum_{j=0}^{\lfloor D^m/2 \rfloor} (-1)^j (2j+1) A_{2j+1}) \cos\lambda \\
&\quad + \ldots + 2^{D^m-1} A_{D^m} \cos^{D^m}\lambda \\
&\overset{(b)}{=} A'_{m,0} + A'_{m,1} \cos\lambda + \ldots + A'_{m,D^m} \cos^{D^m}\lambda & \text{(B.7)}
\end{aligned}$$

The coefficients $A_{m,i}$ and $A'_{m,i}$ are defined by comparing the cosine terms in (a) and (b), respectively, with the corresponding terms in the previous equations.

We have the following relation between $A_{m,i}$s and $h_{m,i}$s:

$$|H_m(0)|^2 = \sum_{i=0}^{D^m} A_{m,i} = (h_{m,0} + \ldots + h_{m,D^m})^2 \tag{B.8}$$

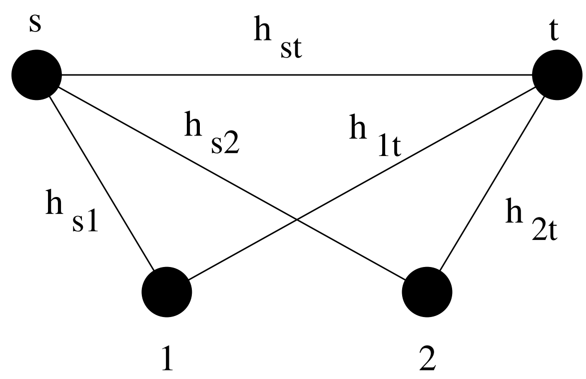

Fig. 7. Relay Network of Type A with M=2.

We can rewrite $\sum_{m=1}^{M} |H_m(\lambda)|^2$ as

$$\begin{aligned}
1 &+ \sum_{m=1}^{M} |H_m(\lambda)|^2 \\
&= (1+\sum_{m=1}^{M} A_{m,0}) + (\sum_{m=1}^{M} A_{m,1})\cos\lambda + \ldots + (\sum_{m=1}^{M} A_{m,D^n})\cos(D^n\lambda) \\
&\overset{(a)}{=} B_0 + B_1\cos\lambda + \ldots + B_{D^n}\cos(D^n\lambda) \qquad \text{(B.9)} \\
&= (1+\sum_{m=1}^{M} A'_{m,0}) + (\sum_{m=1}^{M} A'_{m,1})\cos\lambda + \ldots + (\sum_{m=1}^{M} A'_{m,D^n})\cos^{D^n}\lambda \\
&\overset{(a)}{=} B'_0 + B'_1\cos\lambda + \ldots + B'_{D^n}\cos^{D^n}\lambda \qquad \text{(B.10)}
\end{aligned}$$

where $D^n = \max(D^1, \ldots, D^M)$. The coefficients B_i and B'_i are defined by comparing the cosine terms in (a) and (b) with the corresponding terms in the previous equations.

APPENDIX C

ILLUSTRATION OF THE PROBLEM P2 FOR TWO TYPES OF RELAY NETWORKS

Type A Relay Network: Let us first consider the relay network configuration in Figure 2. This is the simplest instance of the general class of relay networks we consider in this paper. For $M = 2$ relay nodes, the Type A relay network is given in the Figure 7. We illustrate the problem formulation (11) for the network in Figure 7.

The ISI channel between the source and the destination for this network is:

$$\begin{aligned} y[n] &= h_{st}x[n] + (h_{s1}\beta_1 h_{1t} + h_{s2}\beta_2 h_{2t})x[n-1] \\ &\quad +\beta_1 h_{1t} z_1[n-1] + \beta_2 h_{2t} z_2[n-1] + z[n], \end{aligned} \tag{C.1}$$

where

$$0 \leq \beta_1^2 \leq \beta_{1,max}^2 = \frac{P_1}{h_{s1}^2 P_s + \sigma^2}$$
$$0 \leq \beta_2^2 \leq \beta_{2,max}^2 = \frac{P_2}{h_{s2}^2 P_s + \sigma^2}$$

Comparing (C.1) with (7) defines corresponding *modified* channel parameters as follows.

$$\begin{aligned} h_0 &= h_{st}, \quad h_1 = h_{s1}\beta_1 h_{1t} + h_{s2}\beta_2 h_{2t} \\ h_{1,1} &= \beta_1 h_{1t}, \quad h_{2,1} = \beta_2 h_{2t} \end{aligned}$$

Using definitions in (B.2) and (B.9) results in:

$$\begin{aligned} A_0 &= h_0^2 + h_1^2, \quad A_1 = 2h_0 h_1 \\ B_0 &= 1 + h_{1,1}^2 + h_{2,1}^2 \end{aligned}$$

Therefore, we have the following formulation of the problem (11) for the relay network in Figure 7:

$$I_{AF}(P_s) = \max \frac{1}{2\pi} \int_0^{\pi} \log \left[1 + \frac{P_s}{\sigma^2} \frac{A_0 + A_1 \cos\lambda}{B_0} \right] d\lambda \tag{C.2}$$

$$\begin{aligned} \text{subject to} \qquad\qquad 0 &\leq \beta_1^2 \\ 0 &\leq \beta_2^2 \\ \beta_1^2 (h_{s1}^2 P_s + \sigma^2) &\leq P_1 \\ \beta_2^2 (h_{s2}^2 P_s + \sigma^2) &\leq P_2 \end{aligned}$$

Type B Relay Network: Let us next consider the network in Figure 3. For $M = 2$ relay nodes, the Type B relay network is given in the Figure 8. For $M = 2$ relay nodes, the source-destination

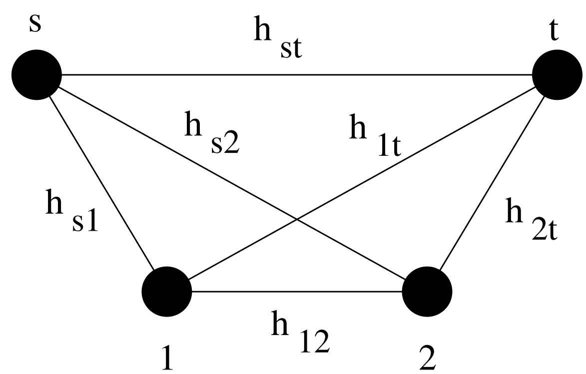

Fig. 8. Relay Network of Type B with M=2.

channel for this relay configuration is:

$$\begin{aligned} y[n] &= h_{st}x[n] + +z[n] \\ &\quad +(h_{s1}\beta_1 h_{1t} + h_{s2}\beta_2 h_{2t})x[n-1] \\ &\quad +(h_{s1}\beta_1 h_{12}\beta_2 h_{2t} + h_{s2}\beta_2 h_{21}\beta_1 h_{1t})x[n-2] \\ &\quad +\beta_1 h_{1t} z_1[n-1] + \beta_1 h_{12}\beta_2 h_{2t} z_1[n-2] \\ &\quad +\beta_2 h_{2t} z_2[n-1] + \beta_2 h_{21}\beta_1 h_{1t} z_2[n-2], \end{aligned} \tag{C.3}$$

where

$$\begin{aligned} 0 \le \beta_1^2 \le \beta_{1,max}^2 &= \frac{P_1}{(h_{s2}^2\beta_2^2 h_{21}^2 + h_{s1}^2)P_s + (\beta_2^2 h_{21}^2 + 1)\sigma^2} \\ 0 \le \beta_2^2 \le \beta_{2,max}^2 &= \frac{P_2}{(h_{s1}^2\beta_1^2 h_{12}^2 + h_{s2}^2)P_s + (\beta_1^2 h_{12}^2 + 1)\sigma^2} \end{aligned}$$

Comparing (C.1) with (7) defines corresponding *modified* channel parameters as follows.

$$h_0 = h_{st}, \quad h_1 = h_{s1}\beta_1 h_{1t} + h_{s2}\beta_2 h_{2t}, \quad h_2 = h_{s1}\beta_1 h_{12}\beta_2 h_{2t} + h_{s2}\beta_2 h_{21}\beta_1 h_{1t}$$

$$h_{1,1} = \beta_1 h_{1t}, \quad h_{1,2} = \beta_1 h_{12}\beta_2 h_{2t}, \quad h_{2,1} = \beta_2 h_{2t}, \quad h_{2,2} = \beta_2 h_{21}\beta_1 h_{1t}$$

Using definitions in (B.2) and (B.9) results in:

$$\begin{aligned} A_0 &= h_0^2 + h_1^2 + h_2^2,\ A_1 = 2(h_0 + h_2)h_1,\ A_2 = 2h_0 h_2 \\ B_0 &= 1 + h_{1,1}^2 + h_{1,2}^2 + h_{2,1}^2 + h_{2,2}^2,\ B_1 = 2h_{1,1}h_{1,2} + 2h_{2,1}h_{2,2} \end{aligned}$$

Therefore, we have the following formulation of the problem (11) for the relay network in Figure 8:

$$I_{AF}(P_s) = \max \frac{1}{2\pi}\int_0^{\pi} \log\left[1+\frac{P_s}{\sigma^2}\frac{\sum_{i=0}^{2} A_i \cos i\lambda}{B_0 + B_1\cos\lambda}\right] d\lambda \tag{C.4}$$

$$\begin{aligned} \text{subject to} \qquad\qquad\qquad\qquad\qquad 0 &\leq \beta_1^2 \\ 0 &\leq \beta_2^2 \\ \beta_1^2((h_{s2}^2\beta_2^2h_{21}^2 + h_{s1}^2)P_s + (\beta_2^2h_{21}^2+1)\sigma^2) &\leq P_1 \\ \beta_2^2((h_{s1}^2\beta_1^2h_{12}^2 + h_{s2}^2)P_s + (\beta_1^2h_{12}^2+1)\sigma^2) &\leq P_2 \end{aligned}$$

Note the nonconvex nature of the constraint set.

APPENDIX D
PROOFS OF LEMMAS AND PROPOSITIONS IN SECTION IV

Proof of Proposition 1: We have from (5) the following nonzero modified channel parameters:

$$h_0 = 1,\ h_1 = \sum_{j=1}^{M} \beta = M\beta$$
$$h_{m,1} = \beta,\ \text{for all } m, 1 \leq m \leq M$$

Therefore, from (B.3), (B.6), and (B.9), we have:

$$\begin{aligned} A_0 + A_1 &= (1+M\beta)^2 \\ B_0 &= 1 + M\beta^2 \end{aligned}$$

Substituting these in the problem (P4) establishes the Proposition.

Proof of Lemma 2: Setting the first-derivative of the objective function equal to zero gives $\beta = 1$ as globally optimal solution.

Also, the objective function is quasiconcave over the domain of optimization. Note that the objective function is nondecreasing for all $\beta < 1$ and nonincreasing for all $\beta > 1$, then its quasiconcavity follows from [21, Proposition 3].

Proof of Proposition 2: We have from (5) the following nonzero modified channel parameters:

$$h_0 = h_{st},\ h_1 = \beta\sum_{i=1}^{M} h_{si}h_{it}$$
$$h_{m,1} = \beta h_{mt},\ \text{for all } m, 1 \leq m \leq M$$

Therefore, from (B.3), (B.6), and (B.9), we have:

$$\begin{aligned} A_0 + A_1 &= (h_{st} + \beta \sum_{i=1}^{M} h_{si} h_{it})^2 \\ B_0 &= 1 + \beta^2 \sum_{i=1}^{M} h_{it}^2 \end{aligned}$$

Substituting these in the problem (P4) establishes the Proposition.

Proof of Lemma 3: Setting the first-derivative of the objective function equal to zero gives $\beta_{opt} = \frac{\sum_{i=1}^{M} h_{si} h_{it}}{h_{st} \sum_{i=1}^{M} h_{it}^2}$ as the globally optimal solution.

Noting that the objective function is nondecreasing for all $\beta < \beta_{opt}$ and nonincreasing for all $\beta > \beta_{opt}$, establishes its quasiconcavity from [21, Proposition 3].

Proof of Proposition 3: We have from (5) the following nonzero modified channel parameters:

$$\begin{aligned} & h_i = \sum_{j=1}^{M-(i-1)} \beta^i = (M-(i-1))\beta^i, 1 \le i \le M, \\ & h_{1,i} = \beta^i, 1 \le i \le M, \\ & h_{2,i} = \beta^1, 1 \le i \le M-1, \\ & \vdots \\ & h_{M,1} = \beta \end{aligned}$$

Therefore, from (B.3) and (B.8), we have:

$$\begin{aligned} |H(0)|^2 &= \left(1 + \beta M \frac{\beta^M - 1}{\beta - 1} - \frac{M\beta^{M+1}}{\beta - 1} + \beta^2 \frac{\beta^M - 1}{(\beta-1)^2}\right)^2 \\ |H_m(0)|^2 &= \beta^2 \left(\frac{\beta^{M-(i-1)} - 1}{\beta - 1}\right)^2, 1 \le m \le M \end{aligned}$$

Substituting these expressions in (14) establishes the proposition.

APPENDIX E
PROOF OF PROPOSITION 4

Broadcast upper bound: For the broadcast cut, we have the following maximization problem:

$$\max_{p_{X_s, X_1, \ldots, X_M}} I(X_s; Y_t, Y_1, \ldots, Y_M | X_1, \ldots, X_M)$$

subject to the power constraints

$$EX_s^2 \le P_s \quad \text{and} \quad EX_i^2 \le P_i, 1 \le i \le M$$

As the destination node knows the channel gains $h_{it}, i \in S_t$, we have

$$I(X_s; Y_t, Y_1, \ldots, Y_M | X_1, \ldots, X_M) = I(X_s; \tilde{Y}_t, \tilde{Y}_{S_s})$$

where $\tilde{Y}_t = Y_t - \sum_{i \in S_t} h_{it} X_i$, $\tilde{Y}_{S_s} = \{\tilde{Y}_i = Y_i - \sum_{j \in \mathcal{N}(i)-s} h_{ji} X_j : i \in S_s\}$. Therefore, the maximum achievable rate of information transfer across the particular broadcast cut C_{BC} of the relay network is given by

$$C_{BC} = \max_{p_{X_s}: EX_s^2 \leq P_s} I(X_s; \tilde{Y}_t, \tilde{Y}_{S_s})$$

The right-hand side can be evaluated using the capacity formula for Gaussian vector channels with fixed transfer function in [23]. We have a channel with scalar input and vector output, therefore corresponding channel matrix has only one singular value given by $a(N_s) = h_{st}^2 + \sum_{i \in S_s} h_{si}^2$. Hence, we have

$$C_{BC} = \log \left[1 + \frac{P_s}{\sigma^2} a(N_s) \right]. \tag{E.1}$$

Multiple-access upper bound: For the multiple-access cut, we have the following maximization problem:

$$\max_{p_{X_s, X_1, \ldots, X_M}} I(X_s, X_1, \ldots, X_M; Y_t)$$

subject to the power constraints

$$EX_s^2 \leq P_s \quad \text{and} \quad EX_i^2 \leq P_i, 1 \leq i \leq M$$

However, given that $Y_t = h_{st} X_s + \sum_{i \in X_{S_t}} h_{it} X_i$, we have

$$I(X_s, X_1, \ldots, X_M; Y_t) = I(X_s, X_{S_t}; Y_t)$$

Therefore the maximization problem above reduces to

$$\max_{p_{X_s, X_{S_t}}} I(X_s, X_{S_t}; Y_t) \tag{E.2}$$

subject to the power constraints

$$EX_s^2 \leq P_s \quad \text{and} \quad EX_i^2 \leq P_i, i \in S_t \tag{E.3}$$

Let us relax this power constraint to

$$EX_s^2 + \sum_{i \in S_t} EX_i^2 \leq P_s + \sum_{i \in S_t} P_i \tag{E.4}$$

The solution of the problem defined by (E.2) and (E.3) cannot be larger than the solution of the problem defined by (E.2) and (E.4). The latter can be evaluated by using the abovementioned result in [23] again. We have a channel with vector input and scalar output, therefore the corresponding channel matrix has only one singular value given by $d(N_t) = h_{st}^2 + \sum_{i \in S_t} h_{it}^2$. Therefore, the maximum achievable rate of information transfer across the particular multiple-access cut C_{MAC} of the relay network is given by

$$C_{MAC} = \log\left[1 + \frac{P_s + \sum_{i \in S_t} P_i}{\sigma^2} d(N_t)\right] \tag{E.5}$$

As the capacity C of the relay network must be smaller than the minimum of C_{BC} and C_{MAC}, this proves the proposition.